Elevating Airline Customer Experience: An Analytical and Predictive Approach to Satisfaction Metrics

Exploratory Data Analysis :

I. Introduction:

This report aims to analyse customer satisfaction levels in the airline industry based on a provided dataset. The report provides a detailed dataset analysis, identifies key findings, and provides recommendations for improving customer satisfaction levels.

II.Executive Summary:

The analysis of the dataset reveals several key findings:

33 rows contained missing values which were deleted, and 1020 rows contained zero values which were also deleted, resulting in a new dataset with a dimension of (9863,23). The overall satisfaction of the data set is 43%, while dissatisfaction and neutrality represent 57% of the observations. The age range of customers is normally distributed with a mean of 40 years old, and satisfaction is higher for customers in the [40-60] age range.

The majority of customers are loyal customers (84%) with an equal distribution of satisfaction, while disloyal customers expressed a high level of dissatisfaction (83%). Business travel is the most common type of travel in the given dataset with a percentage of 69%, and personal travelling customers exhibited a high level of dissatisfaction (91%). The online boarding feature showed a highly significant correlation with customer satisfaction.

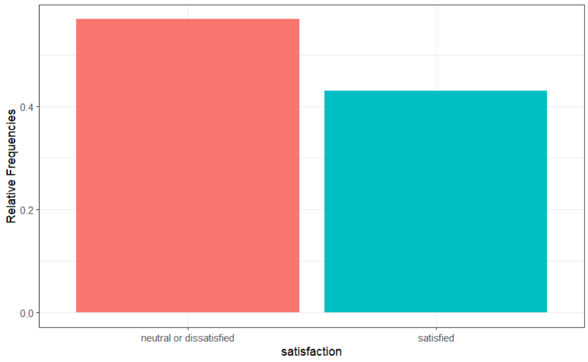
Based on these findings, it is recommended that airlines prioritize improving the online boarding process, particularly for personal travel customers.

III.Data Cleaning: Dealing with Missing and Zero Values:

The summary of the data highlights the existence of 33 missing values (NA's) in the "Arrival.Delay.in.minutes", these relative rows of the missing data were deleted since the feature arrival delay in minutes is crucial to determine the satisfaction of a customer. Additionally, 1020 0 values were noticed in the columns "Inflight.wifi.service", "Ease.of.Online.booking", "Food.and.drink", "Online.boarding", "Inflight.entertainment", "Leg.room.service" and "Cleanliness". Although the 0 values could be interpreted as a typing error of dissatisfaction and replaced with 1 (the lowest value of dissatisfaction), this can create an imbalance in the dataset as well as a biased interpretation if the 0 values do not represent a low value of dissatisfaction. Therefore, the rows with the relative zero values were also dropped, resulting in a new dimension of (9863,23).

Target Variable::

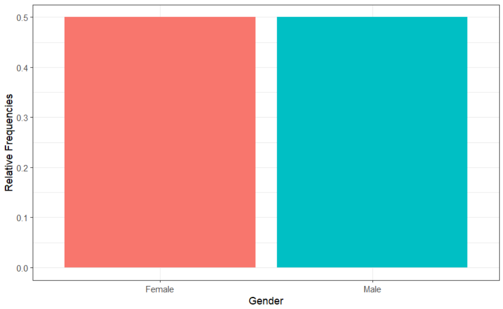
The analysis of the target variable showcases that the overall satisfaction of the dataset is 43%, while dissatisfaction and neutrality represent 57% of the observations.

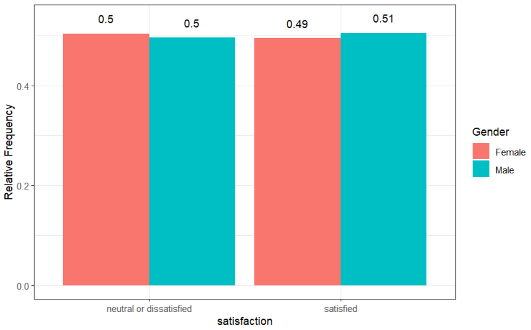


IV.Personal Details:

1.Gender:

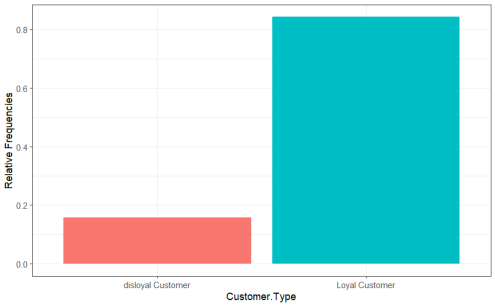
The analysis of gender features of the given dataset suggests a balanced distribution of gender with a 50% occurrence of both genders, and there is also a balanced conditional probability of satisfaction for both genders.

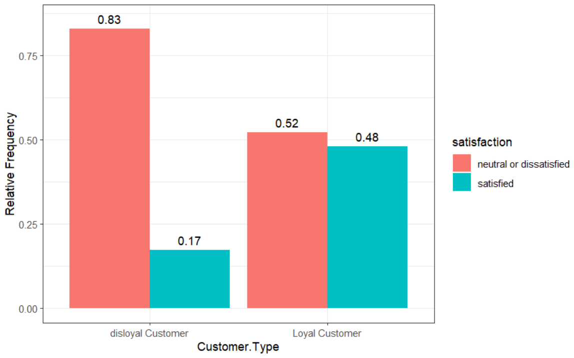




2.Customer Type:

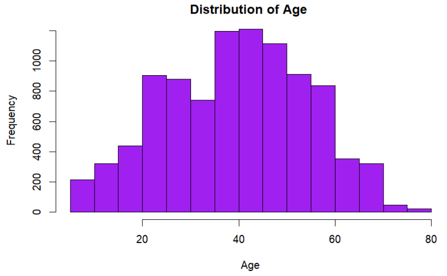
The analysis of customer type highlights that 84% of the population are loyal customers expressing an equal distribution of satisfaction, while disloyal customers expressed a high level of dissatisfaction (83%).

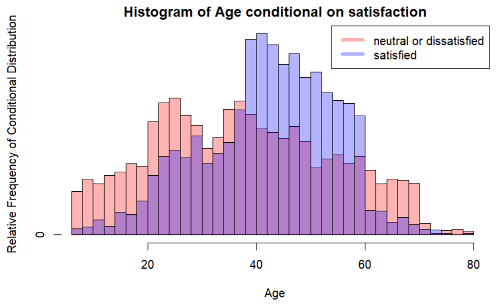


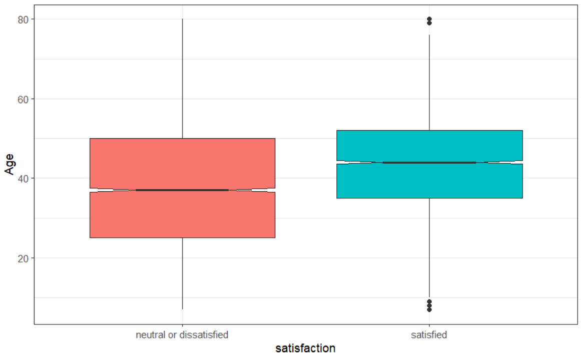


3.Age:

The age feature in the given dataset outlines that the airline customers' mean age is 40 years old following a normal distribution, and this feature also indicates a higher level of satisfaction for the [40-60] age range in contrast with younger ranges [0-40] which expressed a higher level of dissatisfaction. Conditionally, the [60-80] range also showed significant dissatisfaction.

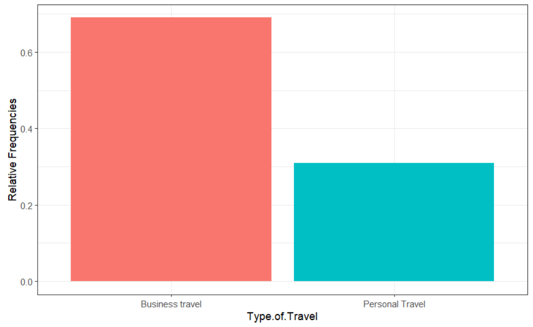


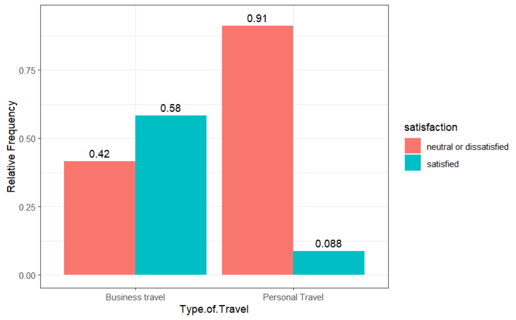




4.Type of Travel:

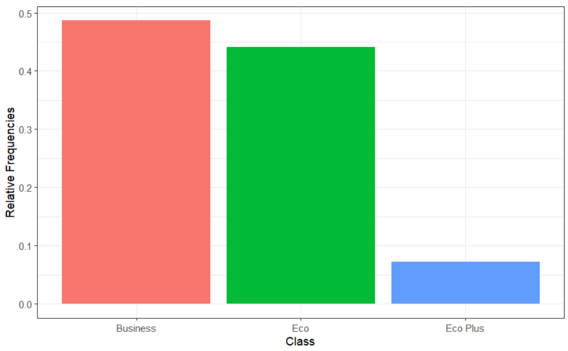
Business travel is the most common type of travel in the given dataset with a percentage of 69%, and personal travelling customers exhibited a high level of dissatisfaction (91%).

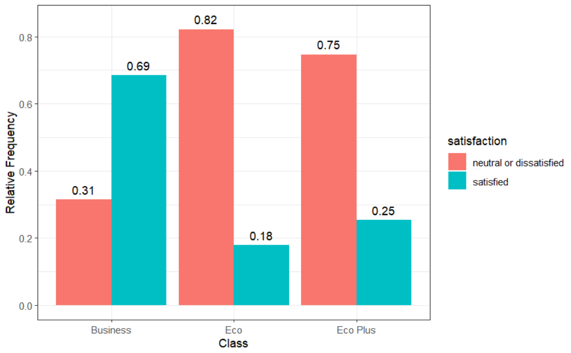




4. Class:

Business travel is the most commun type of travelling in the given data set with a percentage of 69% and 31% for the personal travel type. The personal travelling customers exhibited a high level of dissatisfaction (91%) in contrast the business travel users showcased a good satisfaction level (58%).

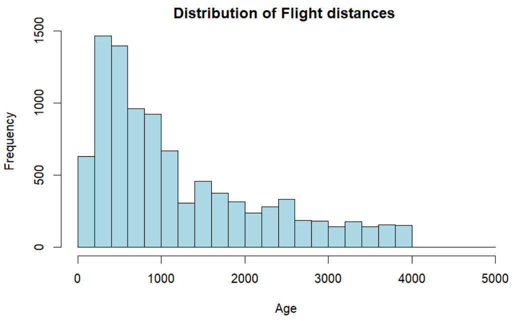


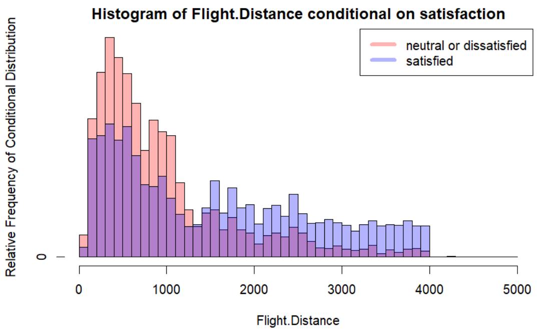


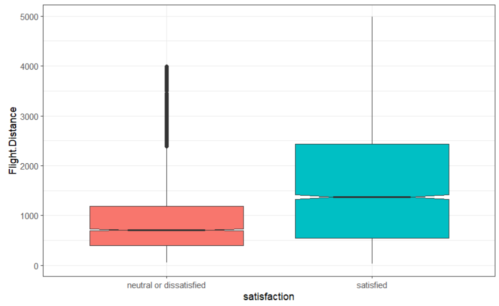
V.Flight Details:

1.Flight Distance:

The Flight Distance feature indicates that most of the flights are in the range of [0-1200] miles. Additionally, the data also shows that customers express a higher degree of satisfaction as the flight distance increases, particularly for distances above 1300 miles. However, for shorter distance flights in the range of [0-1300] miles, customers expressed higher levels of dissatisfaction and neutrality.

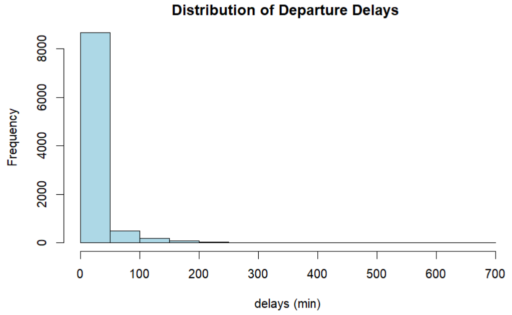


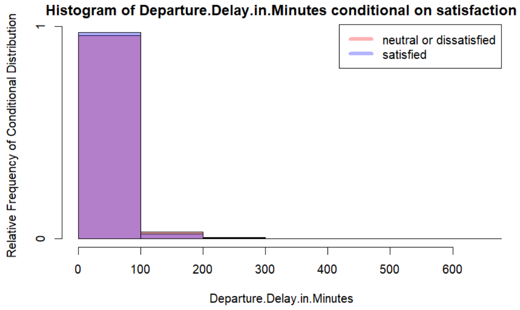


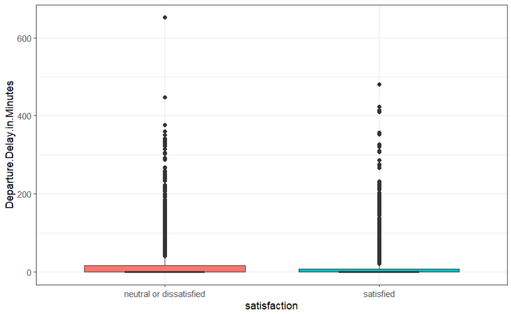


2.Departure Delay:

The Flight Departure Delay feature indicates that a large proportion of flight delays fall in the range of [0-30 minutes], with most of the delays occurring under the [100 minutes] range. Additionally, the data also shows that customers express a higher degree of satisfaction as the departure delays are under the 100-minute range (1 hour 40 minutes). On the other hand, as the delays surpass the 100-minute mark, customers express more dissatisfaction and neutrality.

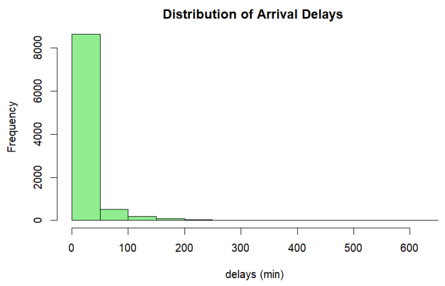


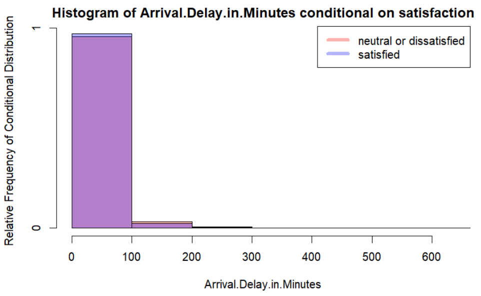


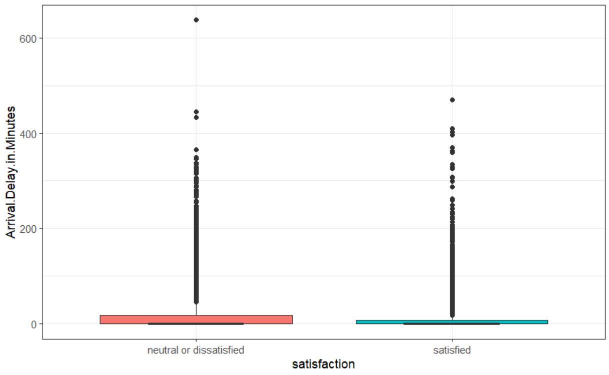


3.Arrival Delay:

The Flight Arrival Delay feature exhibits a similar behavior to the departure delays, where a large proportion of flight delays fall in the range of [0-30 minutes], with most of the delays occurring under the [100 minutes] range. Additionally, the data also shows that customers express a higher degree of satisfaction as the arrival delays are under the 100-minute range (1 hour 40 minutes). On the other hand, as the delays surpass the 100-minute mark, customers express more dissatisfaction and neutrality.



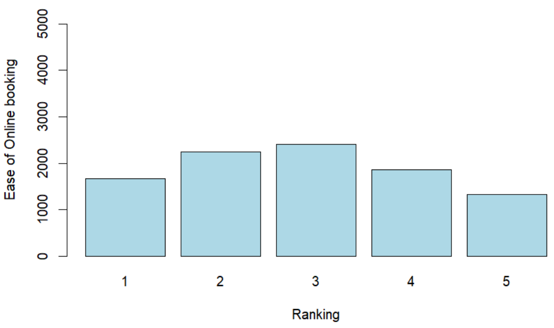




VI.Pre-boarding:

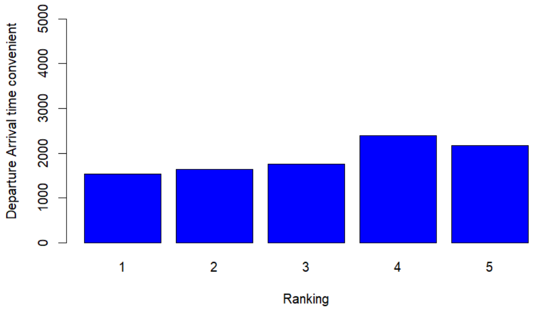
1.Ease of Online Booking:

The Ease of Online Booking feature represents a ranking of satisfaction (1-5). The histogram of the relative ranking is normally distributed, with a slight skew towards rank 2, which highlights a slight dissatisfaction and neutrality.



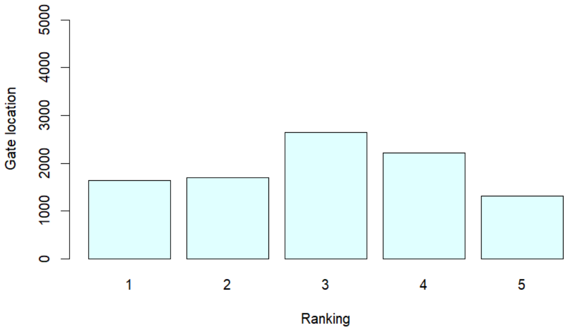
2.Departure Arrival Time Convenient:

The Departure Arrival Time Convenient feature represents a ranking of satisfaction (1-5). The histogram of the relative ranking shows a significant skew towards higher ranks (4-5), which reflects an overall satisfaction with the departure arrival time convenient process.



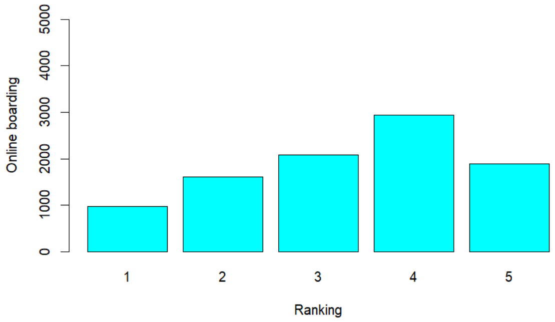
3.Gate Location:

The Gate Location feature represents a ranking of satisfaction (1-5). The histogram of the relative ranking is normally distributed, with a slight skew towards rank 4, which highlights slight satisfaction.



4.Online Boarding:

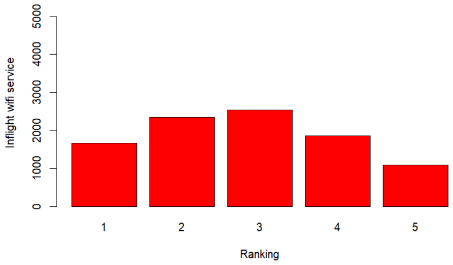
The Online Boarding feature represents a ranking of satisfaction (1-5). The histogram of the relative ranking shows a significant skew towards rank 4, which reflects an overall satisfaction with the Online Boarding process.



VII.On Boarding:

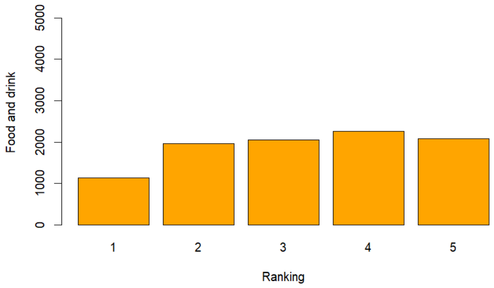
1.Inflight Wifi Service:

The Inflight Wifi Service feature represents a ranking of satisfaction (1-5). The histogram of the relative ranking is normally distributed, with a slight skew towards rank 2, which highlights slight dissatisfaction and neutrality.



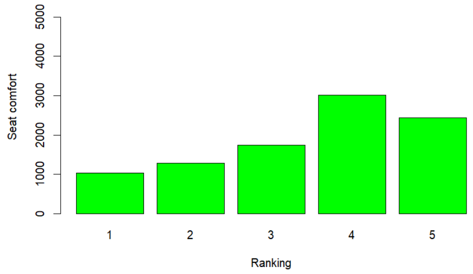
2.Food and Drink:

The Food and Drink feature represents a ranking of satisfaction (1-5). The histogram of the relative ranking shows a significant skew towards higher ranks (3-4-5), which reflects an overall satisfaction with the food and drinks provided while on board.



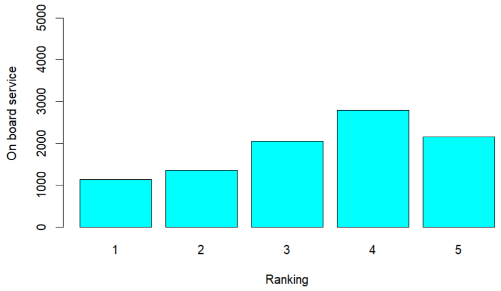
3.Seat Comfort:

The Seat Comfort feature represents a ranking of satisfaction (1-5). The histogram of the relative ranking is normally distributed, with a slight skew towards ranks 4-5, which highlights overall satisfaction.



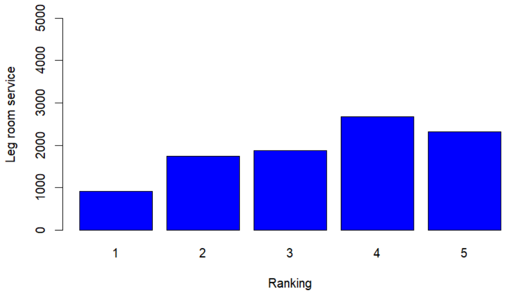
4.On-board Service:

The On-board Service feature represents a ranking of satisfaction (1-5). The histogram of the relative ranking shows a significant skew towards ranks 3 and 4, which reflects an overall satisfaction with the on-board service.



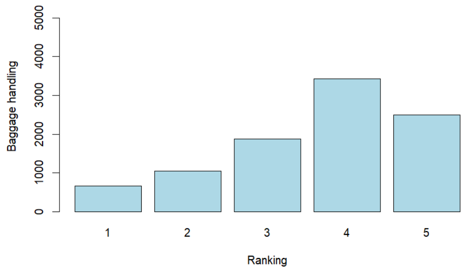
5.Leg Room Service:

The Leg Room Service feature represents a ranking of satisfaction (1-5). The histogram of the relative ranking shows a significant skew towards ranks 4 and 5, which reflects an overall satisfaction with the leg room service.



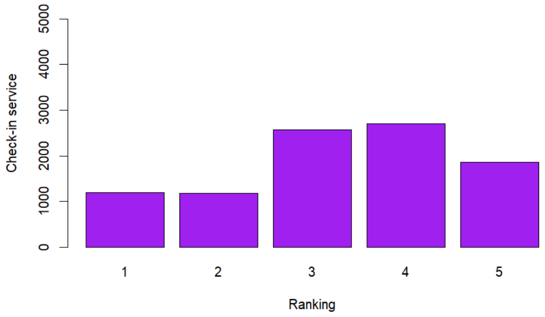
6.Baggage Handling:

The Baggage handling feature represents a ranking of satisfaction (1-5), and the histogram of the relative ranking shows a significant skew towards the ranks 4 and 5, which reflects an overall satisfaction with the baggage handling process. This is an important feature as baggage handling is a critical aspect of the overall travel experience for customers.



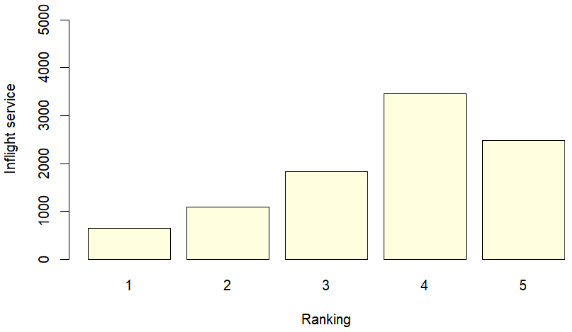
7.Check-in Service:

The Check-in service feature represents a ranking of satisfaction (1-5), and the histogram of the relative ranking shows a significant skew towards the ranks 3 and 4, which reflects an overall satisfaction with the check-in service process. However, there is still room for improvement as there are some customers who expressed dissatisfaction with this aspect of their travel experience.



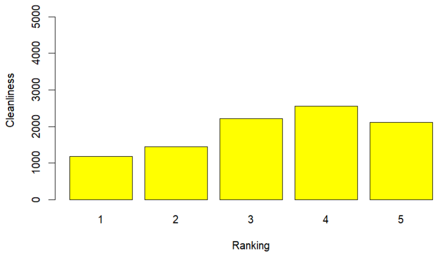
8.Inflight Service:

The Inflight service feature represents a ranking of satisfaction (1-5), and the histogram of the relative ranking shows a significant skew towards the ranks 4 and 5, which reflects an overall satisfaction with the inflight service. This is an important aspect of the travel experience as customers spend a significant amount of time on the flight and expect a high level of service.



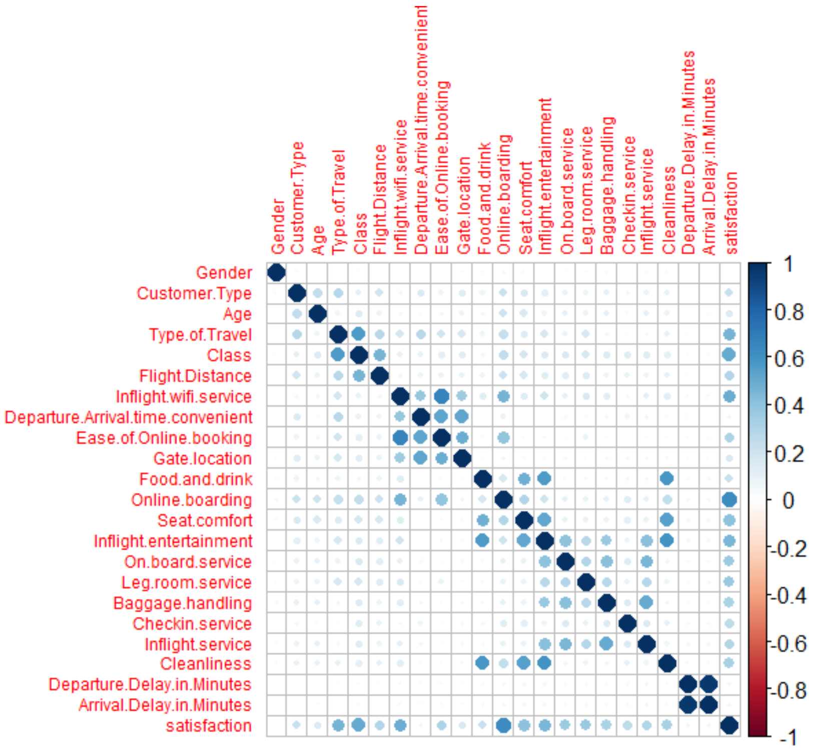
9.Cleanliness:

The Cleanliness feature represents a ranking of satisfaction (1-5), and the histogram of the relative ranking shows a significant skew towards the ranks 3, 4, and 5, which reflects an overall satisfaction with the cleanliness of the flights. This is another critical aspect of the travel experience as customers expect a clean and hygienic environment during their flight.



VIII.Correlation:

The analysis of the correlation between the different features and the target variable (satisfaction) shows that gender, customer type, and age have a low correlation with the target variable. However, Class and customer type have shown a significant correlation with the target variable (satisfaction). The pre-boarding features (Ease of online booking, Departure Arrival time convenient, Gate location, and Online boarding) have also shown a significant correlation with the target variable (satisfaction), especially the online boarding feature (highly correlated with satisfaction). The flight details have shown an insignificant correlation with the target variable except for the flight distance feature which is positively correlated with the target variable. Onboarding features (Inflight wifi service, Food and drink, Seat comfort, Inflight entertainment, Onboard service, Leg room service, Baggage handling, Check-in service, Inflight service, and Cleanliness) have shown an insignificant correlation with the target variable (satisfaction). We also notice a significant correlation between the in-categories features (onboarding and pre-boarding).

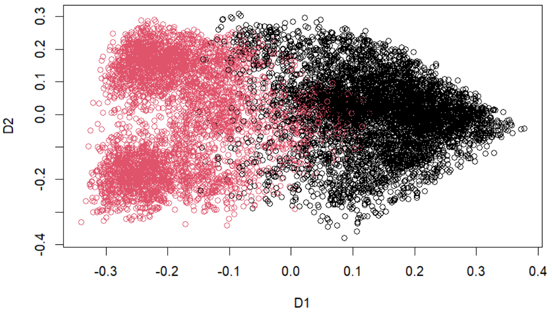


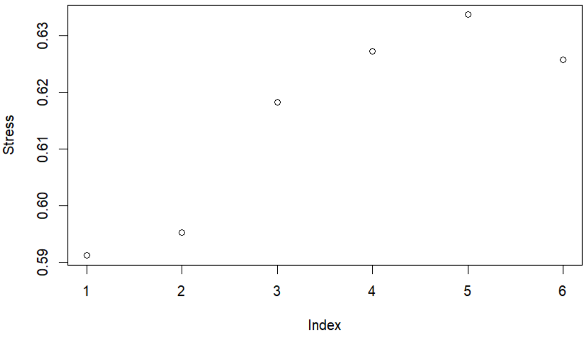
IX.Dimensionality Reduction:

The high correlation between some of the categorical features leads to increased complexity in the data, which can be difficult to interpret. Therefore, a dimensionality reduction technique was applied to the data using Multi-Dimensional Scaling (MDS).

MDS is a statistical technique used to visualize the similarity or dissimilarity between data points in a lower-dimensional space. In this case, MDS was used to create a two-dimensional representation of the categorical features in the dataset. The Gower distance metric was used to calculate the distances between the data points, which considers the difference in data types and scales them accordingly. The stress method was used to determine the optimal number of dimensions to use in the MDS analysis. The stress scores were calculated for different numbers of dimensions, and the best solution was found to be k=1 and k=2, with stress scores of 0.591 and 0.596, respectively. The choice of k=2 was made to allow for a two-dimensional visualization of the data that can be easily interpreted and visualized.

The resulting two-dimensional plot shows a separation between satisfied (black dots), dissatisfied and neutral (red dots) customers. The satisfied customers are clustered towards the top of the plot, while the dissatisfied customers are clustered towards the bottom. There is also some separation between the loyal and disloyal customers, with the loyal customers clustered towards the left of the plot, and the disloyal customers towards the right.

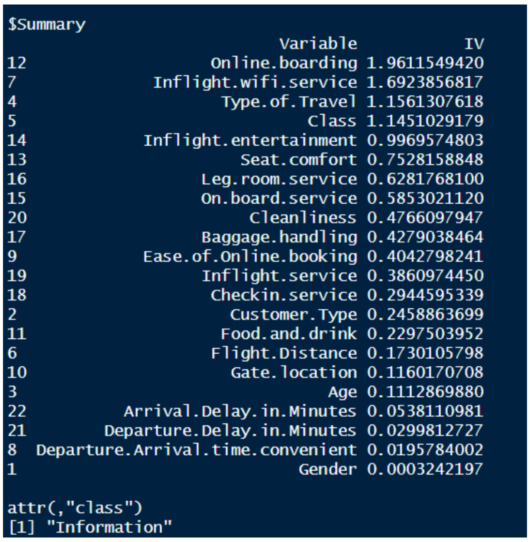




X.Information Value:

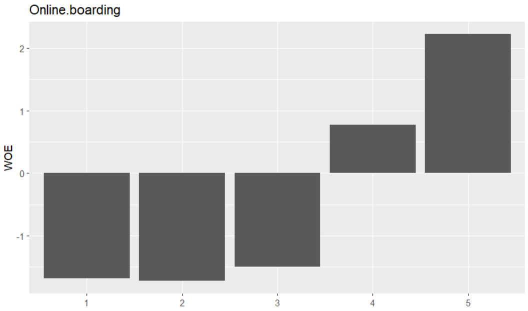
The Information Value (IV) indicates the degree to which a variable is associated with the target variable (in this case, customer satisfaction). The IV analysis was performed on all the features in the dataset, and the results show that the most important features for predicting customer satisfaction are "Online Boarding", "Inflight Wifi Service", "Type of Travel", and "Class". These features have high IV scores, indicating that they are strongly associated with customer satisfaction. Other important features for predicting customer satisfaction include "Inflight Entertainment", "Seat Comfort", "Leg Room Service", and "Onboard Service". These features have moderate IV scores, indicating that they are moderately associated with customer satisfaction. The least important features for predicting customer satisfaction include "Gender", "Departure Arrival Time Convenient", "Arrival Delay in Minutes", and "Departure Delay in Minutes". These features have very low IV scores, indicating that they are not strongly associated with customer satisfaction.

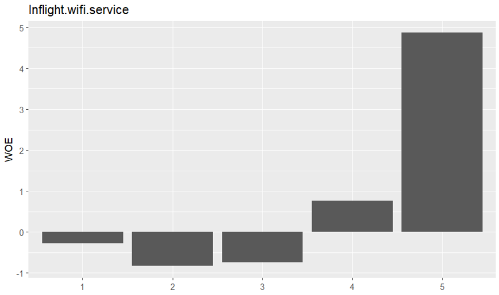
Overall, the IV analysis confirms the results of the correlation analysis, highlighting the importance of certain features in predicting customer satisfaction.

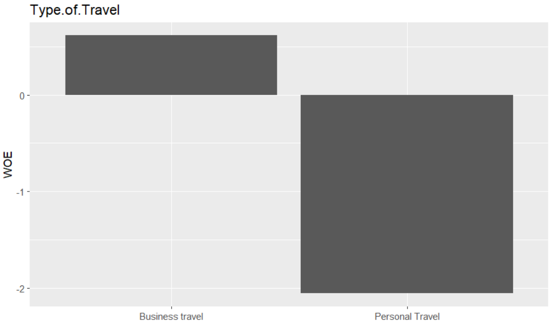


```
$Summary
                            Variable           IV
12                   Online.boarding 1.9611549420
7              Inflight.wifi.service 1.6923856817
4                     Type.of.Travel 1.1561307618
5                              Class 1.1451029179
14            Inflight.entertainment 0.9969574803
13                      Seat.comfort 0.7528158848
16                  Leg.room.service 0.6281768100
15                  On.board.service 0.5853021120
20                       Cleanliness 0.4766097947
17                  Baggage.handling 0.4279038464
9             Ease.of.Online.booking 0.4042798241
19                  Inflight.service 0.3860974450
18                   Checkin.service 0.2944595339
2                      Customer.Type 0.2458863699
11                    Food.and.drink 0.2297503952
6                    Flight.Distance 0.1730105798
10                     Gate.location 0.1160170708
3                                Age 0.1112869880
22          Arrival.Delay.in.Minutes 0.0538110981
21        Departure.Delay.in.Minutes 0.0299812727
8  Departure.Arrival.time.convenient 0.0195784002
1                             Gender 0.0003242197

attr(,"class")
[1] "Information"
```







XI.Conclusion:

In conclusion, the analysis of the airline passenger satisfaction dataset has highlighted several key findings. Firstly, we identified missing values and 0 values that were removed from the data set to avoid biased interpretation. Secondly, we identified the overall satisfaction level of the data set to be at 43%, indicating that there is room for improvement.

We also analyzed the impact of personal details, flight details, pre-boarding, on-boarding, and correlation on passenger satisfaction. Our analysis found that the ease of online booking, departure arrival time convenient, gate location, online boarding, flight distance, type of travel, and customer type had a significant impact on passenger satisfaction.

Moreover, we used multi-dimensional scaling (MDS) to visualize the similarity between data points and identify groups of features that are positively correlated with each other, which can help reduce the complexity of the data.

Finally, we applied the information value (IV) method to rank the importance of each feature on the target variable. The results were consistent with our previous analysis, highlighting the same important features.

Based on these findings, we recommend that airlines focus on improving the pre-boarding and on-boarding experiences, such as online booking, departure arrival time convenient, gate location, and online boarding. Additionally, airlines should consider the type of travel and customer type when designing their services to ensure that their loyal customers are satisfied. Finally, airlines should consider the importance of each feature when allocating their resources for improvement, which can be guided by the IV score.

Overall, this report provides insights into the factors that affect airline passenger satisfaction and provides recommendations to improve the passenger experience, which can lead to increased customer loyalty and profitability for airlines.

Executive Summary :

The analysis and modelling conducted on the "airlinesData120.csv" dataset involved exploratory data analysis, dimension reduction using Multidimensional Scaling (MDS), and the development of predictive models for customer satisfaction in the airline industry, the dataset was pre-processed by handling missing and zero values, resulting in a dataset with dimensions (9863, 23).

Based on the findings from the initial analysis, variables such as Inflight Wifi Service, Ease of Online Booking, Food and Drink, Online Boarding, Inflight Entertainment, Leg Room Service, and Cleanliness were identified as key contributors to predicting customer satisfaction. These variables, representing different aspects of the customer experience, significantly impact overall satisfaction levels. Multidimensional Scaling (MDS) was used to visualize the relationship between customer satisfaction and other variables, the benchmark models were developed using logistic regression, KNN, and decision tree algorithms with all features as response variables. These models were evaluated on training and testing datasets, considering various performance metrics such as accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV), to further improve the models' accuracy, fine-tuning was performed. Stepwise logistic regression was used to refine the model by iteratively adding and removing variables based on their significance. KNN models were fine-tuned by adjusting the training parameters, such as the number of neighbours (k), using repeated cross-validation. Decision tree models were enhanced using bagging, an ensemble learning technique that combines multiple models trained on subsets of the original dataset.

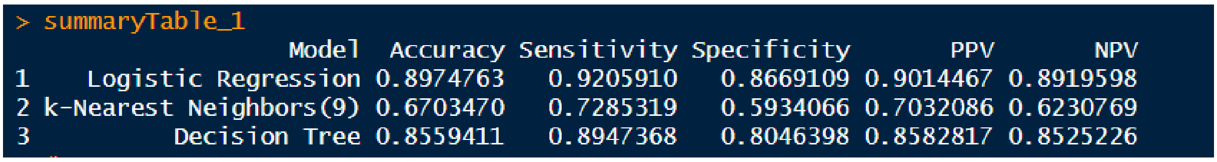
Evaluation of the benchmark and fine-tuned models was conducted, considering accuracy, sensitivity, specificity, PPV, NPV, and the area under the ROC curve (AUC). The AUC provides an overall measure of the model’s discriminatory power. The logistic regression model (stepwise) and decision tree model (with bagging) exhibited the highest accuracy and AUC, making them the recommended models for predicting customer satisfaction.

In conclusion, the logistic regression (stepwise) and decision tree (with bagging) models demonstrated superior performance in predicting customer satisfaction in the airline industry. They achieved high accuracy, sensitivity, and specificity, while also exhibiting a balanced trade-off between correct identification of satisfaction and dissatisfaction and meeting the stakeholder’s requirements (95% sensitivity, and 90% specificity). The logistic regression model offers interpretability through its coefficients, aiding in understanding the factors influencing customer satisfaction.

Predictive Analysis & Modelling :

Benchmark models :

The modelling process proceeds by building benchmark models using logistic regression, k-nearest neighbours, and decision tree algorithms(by taking all the features as a response variable). The dataset is split into training and testing sets, each model is trained using cross-validation on the training data and evaluated on the testing data, the models' formulas consider all the features resulting from the pre-processing. The accuracies, sensitivity, specificity, positive predictive value, and negative predictive value of each model are calculated and presented in summary tables.[Figure 4]



```
> summaryTable_1
                 Model  Accuracy Sensitivity Specificity       PPV       NPV
1    Logistic Regression 0.8974763   0.9205910   0.8669109 0.9014467 0.8919598
2 k-Nearest Neighbors(9) 0.6703470   0.7285319   0.5934066 0.7032086 0.6230769
3          Decision Tree 0.8559411   0.8947368   0.8046398 0.8582817 0.8525226
```

Figure 4 : summary results of benchmark models

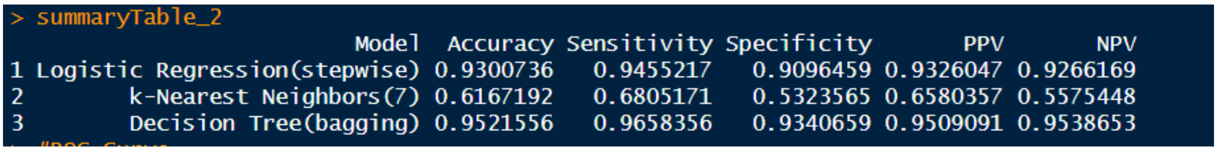
Fine-tuned models :

In the second part of modelling, the benchmark models are finely tuned in order to increase the accuracy of prediction using various approaches, namely stepwise logistic regression, fine-tuned k-nearest neighbours (knn), and decision trees with bagging.

For stepwise logistic regression, the stepwise selection process is utilized to refine the model. The train() function with the "glmStepAIC" method is employed by iteratively adding and removing variables based on their significance, which implements stepwise model selection using the Akaike information criterion (AIC). This approach automatically selects the most relevant predictors from the dataset, considering their impact on the model's accuracy. the stepwise logistic regression model fine-tunes its predictor set to improve the overall model fit.

In the case of knn, fine-tuning is performed on the model's training parameters by adjusting the control of training in order to obtain the best-performing k (k=9)by varying the k value and utilizing repeated cross-validation, the knn model is fine-tuned to find the optimal number of neighbours, with 10 folds and 5 repetitions. Additionally, knnGrid is defined as a grid object, specifying the range of k values from 1 to 10.

For the decision tree algorithm, the "treebag" method is used to build a decision tree model with bagging, short for bootstrap aggregating which is an ensemble learning technique that combines multiple models to make predictions, aiming to reduce variance and improve the overall performance of machine learning algorithms. The main idea behind bagging is to create multiple subsets of the original dataset through a process called bootstrap sampling. In bootstrap sampling, each subset is created by randomly selecting observations from the original dataset with replacement, this means that some observations may appear multiple times in a subset, while others may be left out. This process creates diversity among the subsets and helps to improve the overall performance of the ensemble model by reducing overfitting, increasing robustness against noise and outliers, and enhancing the model's generalization ability. It is particularly effective when the base models are weak learners that have low bias and high variance.

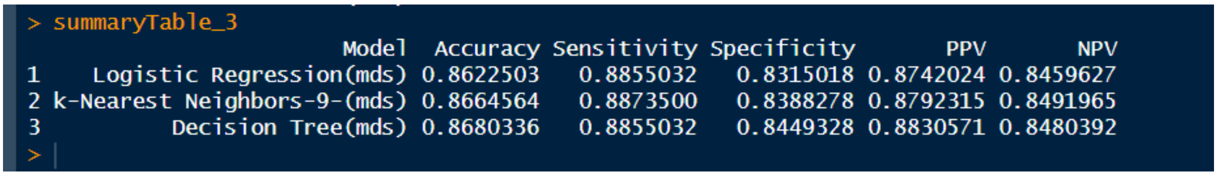


```
> summaryTable_2
                        Model  Accuracy Sensitivity Specificity       PPV       NPV
1 Logistic Regression(stepwise) 0.9300736   0.9455217   0.9096459 0.9326047 0.9266169
2         k-Nearest Neighbors(7) 0.6167192   0.6805171   0.5323565 0.6580357 0.5575448
3         Decision Tree(bagging) 0.9521556   0.9658356   0.9340659 0.9509091 0.9538653
```

Figure 5 : summary results of the fine-tuning

MDS models :

After performing Multidimensional Scaling (MDS) on the pre-processed data to visualize the relationship between customer satisfaction and other variables. The dissimilarity matrix is computed using daisy() with the "gower" metric, and cmdscale() is applied to obtain a 2-dimensional representation of the data. Furthermore, the model-building and evaluation process is repeated using the MDS-transformed data and Logistic regression, k-nearest neighbours, and decision tree models are trained and tested on the transformed dataset. The accuracies, sensitivity, specificity, PPV, NPV values are recorded and presented in summary tables[Figure 6], where the accuracy metric provides an overall measure of the models' correctness in predicting customer satisfaction. Sensitivity captures the model's ability to correctly identify dissatisfied customers, while specificity measures its ability to avoid misclassifying satisfied customers as dissatisfied.



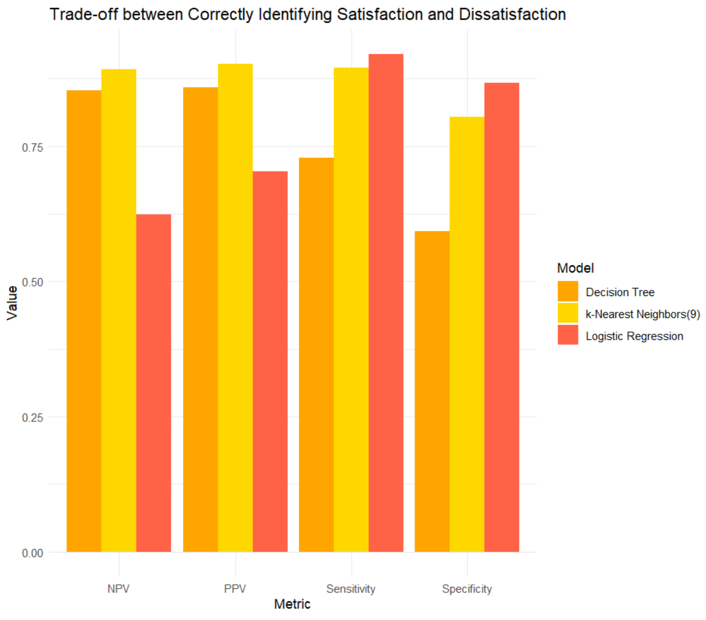
```
> summaryTable_3
                       Model  Accuracy Sensitivity Specificity       PPV       NPV
1    Logistic Regression(mds) 0.8622503   0.8855032   0.8315018 0.8742024 0.8459627
2 k-Nearest Neighbors-9-(mds) 0.8664564   0.8873500   0.8388278 0.8792315 0.8491965
3          Decision Tree(mds) 0.8680336   0.8855032   0.8449328 0.8830571 0.8480392
> 
```

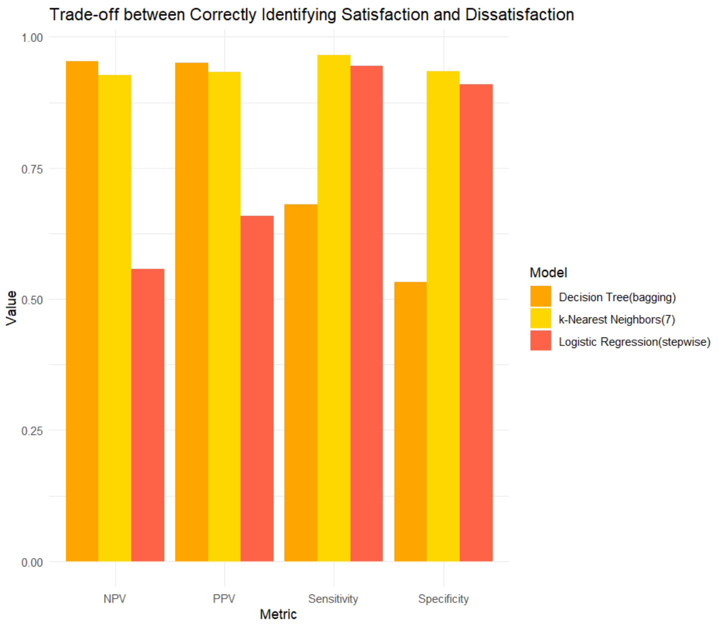
Figure 6: summary results of MDS models

Results :

The trade-off between correctly identifying satisfaction and dissatisfaction is a critical aspect of classification models. This trade-off entails finding an optimal classification threshold that balances between sensitivity (correctly identifying dissatisfied customers) and specificity (avoiding misclassification of satisfied customers as dissatisfied), visualizing this trade-off is achieved through the usage of a receiver operating characteristic (ROC) curve and histograms[Figure 7]. The ROC curve plots the true positive rate against the false positive rate at various classification thresholds, providing insights into the model's performance across different threshold values. A model with a higher ROC curve (closer to the top-left corner) indicates superior performance in correctly classifying both satisfied and dissatisfied customers.

In our analysis, we construct and evaluate ROC curves for each classification model, namely Logistic Regression, k-Nearest Neighbors, and Decision Tree[Figure 8].





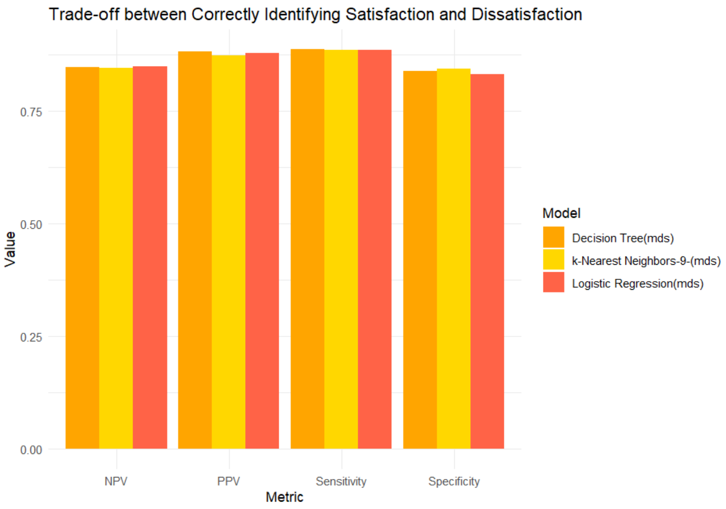


Figure 7 : Trade off between correctly identified satisfied vs dissatisfied (benchmark model-fine-tuned models-MDS models)

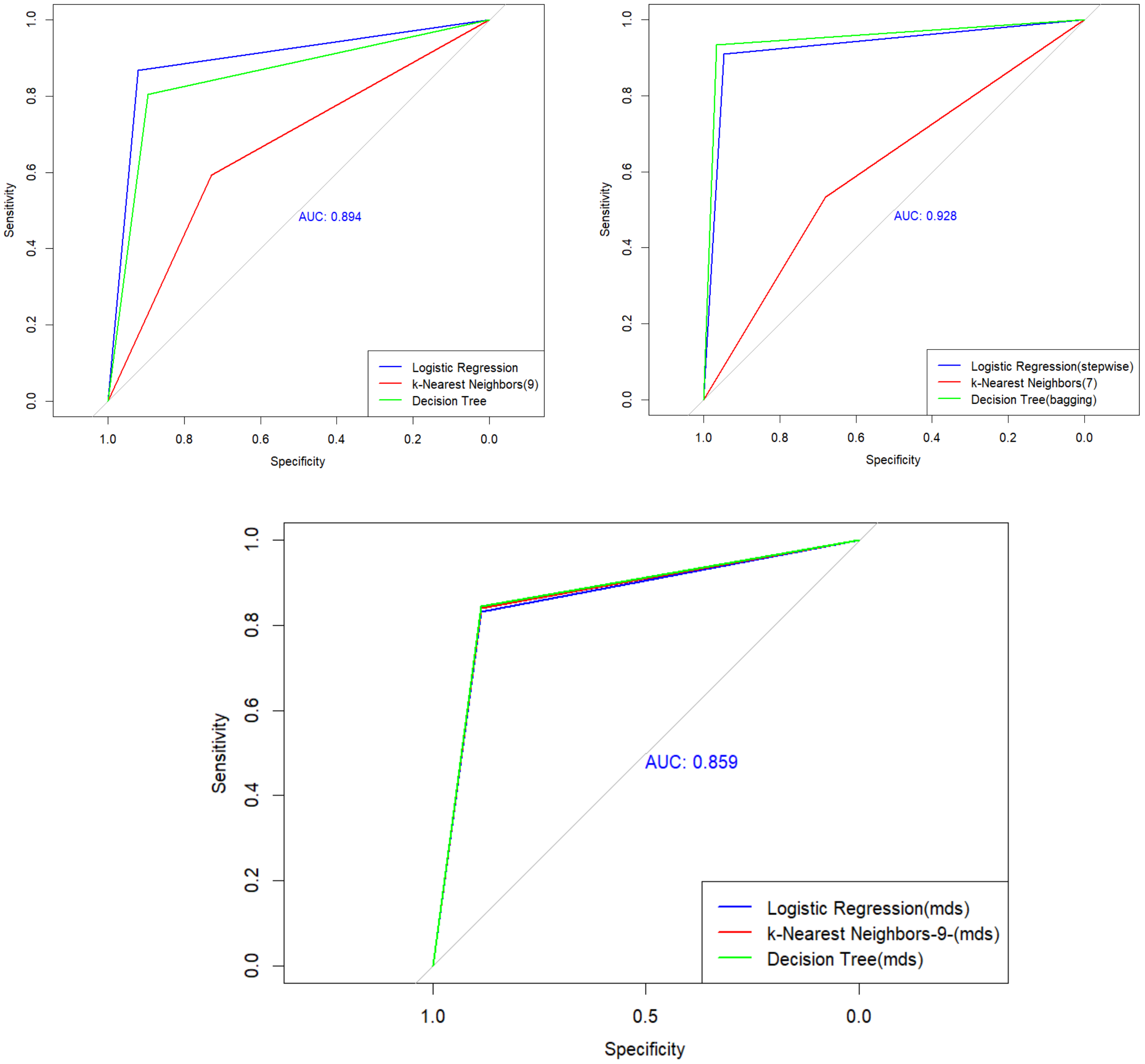
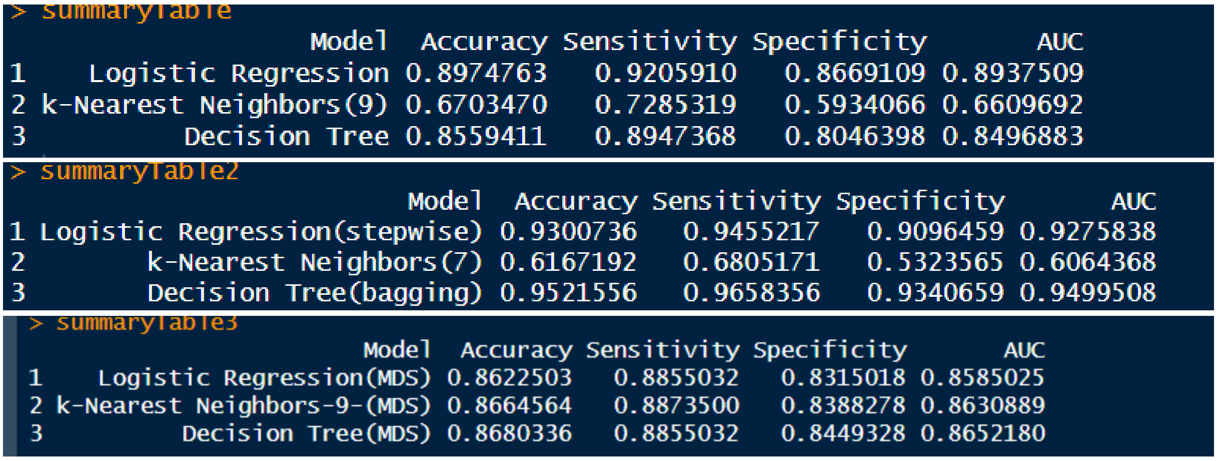


Figure 8 : ROC curves (benchmark model-fine-tuned models-MDS models)

After evaluating the performance of each classification model and considering the trade-off between correct identification of satisfaction and dissatisfaction, the Logistic Regression (stepwise) and decision tree (Bagging) models are recommended as the most appropriate for predicting customer satisfaction, the models demonstrate the highest accuracies compared to the other models, indicating its ability to make correct predictions and meeting the stakeholder's criterias. overall, the models also exhibited high sensitivity, meaning it can accurately identify dissatisfied customers and a reasonable level of specificity, ensuring that the misclassification of satisfied customers as dissatisfied is minimized. The area under the ROC curve (AUC) for the Logistic Regression and Decision tree model is the highest among the models considered. This implies that the model has a better overall discriminatory power and performs well across a range of classification thresholds. It is also important to note that Logistic Regression models offer interpretability by providing coefficients that represent the influence of each variable on the target variable. This interpretability can assist in understanding the factors that contribute to customer satisfaction.

Considering these factors, the Logistic Regression model emerges as the most suitable choice for predicting customer satisfaction in the airline industry. However, it is essential to continue monitoring the model's performance and make adjustments as needed to ensure its accuracy and relevance in real-world deployment scenarios.



```
> summaryTable
                  Model  Accuracy Sensitivity Specificity       AUC
1   Logistic Regression 0.8974763   0.9205910   0.8669109 0.8937509
2 k-Nearest Neighbors(9) 0.6703470   0.7285319   0.5934066 0.6609692
3         Decision Tree 0.8559411   0.8947368   0.8046398 0.8496883
> summaryTable2
                         Model  Accuracy Sensitivity Specificity       AUC
1 Logistic Regression(stepwise) 0.9300736   0.9455217   0.9096459 0.9275838
2        k-Nearest Neighbors(7) 0.6167192   0.6805171   0.5323565 0.6064368
3        Decision Tree(bagging) 0.9521556   0.9658356   0.9340659 0.9499508
> summaryTable3
                      Model  Accuracy Sensitivity Specificity       AUC
1   Logistic Regression(MDS) 0.8622503   0.8855032   0.8315018 0.8585025
2 k-Nearest Neighbors-9-(MDS) 0.8664564   0.8873500   0.8388278 0.8630889
3         Decision Tree(MDS) 0.8680336   0.8855032   0.8449328 0.8652180
```

Figure 9 : summary table of all the models discussed

Conclusions:

In conclusion, the Logistic Regression and decision tree (with bagging) models emerge as the optimal choice for predicting customer satisfaction in the airline industry. Its superior performance, a balanced trade-off between correct identification of satisfaction and dissatisfaction, and ability to incorporate significant variables make it the most suitable model. However, it is crucial to continuously monitor and evaluate the model's performance in real-world scenarios. Ongoing adjustments and improvements ensure that the model remains accurate and relevant. By leveraging the insights gained from this report, marketing managers in the airline industry can make informed decisions and improve customer satisfaction based on a data-driven approach.